

Effects of associated kernels in nonparametric multiple regressions

Sobom M. Somé^{a,*}, Célestin C. Kokonendji^a

^a*University of Franche-Comté, Besançon, France*

Abstract

Associated kernels have been introduced to improve the classical continuous kernels for smoothing any functional on several kinds of supports such as bounded continuous and discrete sets. This work deals with the effects of combined associated kernels on nonparametric multiple regression functions. Using the Nadaraya-Watson estimator with optimal bandwidth matrices selected by cross-validation procedure, different behaviours of multiple regression estimations are pointed out according the type of multivariate associated kernels with correlation or not. Through simulation studies, there are no effect of correlation structures for the continuous regression functions and also for the associated continuous kernels; however, there exist really effects of the choice of multivariate associated kernels following the support of the multiple regression functions bounded continuous or discrete. Applications are made on two real datasets.

Keywords: Bandwidth matrix, continuous associated kernel, correlation structure, cross-validation, discrete associated kernel, Nadaraya-Watson estimator.

Mathematics Subject Classification 2010: 62G05(08); 62H12

Short Running Title: Associated kernels in multiple regressions

1. Introduction

Considering the relation between a response variable Y and a d -vector ($d \geq 1$) of explanatory variables \mathbf{x} given by

$$Y = m(\mathbf{x}) + \epsilon, \quad (1.1)$$

**Corresponding author:* Université de Franche-Comté, UFR Sciences et Techniques, Laboratoire de Mathématiques de Besançon – UMR 6623 CNRS-UFC, 16 route de Gray, 25030 Besançon cedex, France; Tel. +33 381 666 398; Fax +33 381 666 623.

Email addresses: sobom.some@univ-fcomte.fr (Sobom M. Somé), celestin.kokonendji@univ-fcomte.fr (Célestin C. Kokonendji)

where m is the unknown regression function from $\mathbb{T}_d \subseteq \mathbb{R}^d$ to \mathbb{R} and ϵ the disturbance term with null mean and finite variance. Let $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ be a sequence of independent and identically distributed (iid) random vectors on $\mathbb{T}_d \times \mathbb{R} (\subseteq \mathbb{R}^{d+1})$ with $m(\mathbf{x}) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x})$ of (1.1). The [Nadaraya \(1964\)](#) and [Watson \(1964\)](#) estimator \widehat{m}_n of m , using continuous classical (symmetric) kernels is

$$\widehat{m}_n(\mathbf{x}; K, \mathbf{H}) = \sum_{i=1}^n \frac{Y_i K\{\mathbf{H}^{-1}(\mathbf{x} - \mathbf{X}_i)\}}{\sum_{i=1}^n K\{\mathbf{H}^{-1}(\mathbf{x} - \mathbf{X}_i)\}} = \widehat{m}_n(\mathbf{x}; \mathbf{H}), \quad \forall \mathbf{x} \in \mathbb{T}_d := \mathbb{R}^d, \quad (1.2)$$

where \mathbf{H} is the symmetric and positive definite bandwidth matrix of dimension $d \times d$ and the function $K(\cdot)$ is the multivariate kernel assumed to be spherically symmetric probability density function. Since the choice of the kernel K is not important in classical case, we use the common notation $\widehat{m}_n(\mathbf{x}; \mathbf{H})$ for classical kernel regression. The multivariate classical kernel (e.g. Gaussian) suits only for regression functions on unbounded supports (i.e. \mathbb{R}^d); see also [Scott \(1992\)](#). [Racine and Li \(2004\)](#) proposed product of kernels composed by univariate Gaussian kernels for continuous variables and [Aitchison and Aitken \(1976\)](#) kernels for categorical variables; see also [Hayfield and Racine \(2007\)](#) for some implementations and uses of these multiple kernels. Notice that the use of symmetric kernels gives weights outside variables with unbounded supports. In the univariate continuous case, [Chen \(1999, 2000ab\)](#) is one of the pioneers who has proposed asymmetric kernels (i.e. beta and gamma) which supports coincide with those of the functions to be estimated. [Zhang \(2010\)](#) and [Zhang and Karunamuni \(2010\)](#) studied the performance of these beta and gamma kernel estimators at the boundaries in comparison with those of the classical kernels. Recently, [Libengué \(2013\)](#) investigated several families of these univariate continuous kernels that he called univariate associated kernels; see also [Kokonendji et al. \(2007\)](#), [Kokonendji and Senga Kiéssé \(2011\)](#), [Zougab et al. \(2012\)](#) and [Wansouwé et al. \(2014\)](#) for univariate discrete situations. A continuous multivariate version of these associated kernels have been studied by [Kokonendji and Somé \(2015\)](#) for density estimation.

The main goal of this work is to consider multivariate associated kernels and then to investigate the importance of their choice in multiple regression. These associated kernels are appropriated for both continuous and count explanatory variables. In fact, in order to estimate the regression function m in (1.1), we propose multiple (or product of) associated kernels composed by univariate discrete associated kernels (e.g. binomial, discrete triangular) and continuous ones (e.g. beta, Epanechnikov). We will also use a bivariate beta kernel with correlation structure. Another motivation of this work is to investigate the effect of correlation structure for explanatory variables in continuous regression estimation. These associated kernels suit for this situation of mixing axes as they fully respect the support of each explanatory variable. In other words, we will measure the effect of type of associated kernels, denoted κ , in multiple regression by simulations and applications.

The rest of the paper is organized as follows. Section 2 gives a general definition of multivariate associated kernels which includes the continuous classical symmetric and the multiple composed by univariate discrete and continuous. For each definition, the corresponding kernel regression appropriated for both continuous and discrete explanatory variables are given. In Section 3, we explore the importance of the choice of appropriated associated kernels according to the support of the variables through simulations studies and real data analysis. Finally, summary and final remarks are drawn in Section 4.

2. Multiple regression by associated kernels

2.1. Definition

In order to include both discrete and continuous regressors, we assume \mathbb{T}_d is any subset of \mathbb{R}^d . More precisely, for $j = 1, \dots, n$, let us consider on $\mathbb{T}_d = \otimes_{j=1}^d \mathbb{T}_1^{[j]}$ the measure $\nu = \nu_1 \otimes \dots \otimes \nu_d$ where ν_j is a Lesbesgue or count measure on the corresponding univariate support $\mathbb{T}_1^{[j]}$. Under these assumptions, the associated kernel $K_{\mathbf{x}, \mathbf{H}}(\cdot)$ which replaces the classical kernel $K(\cdot)$ of (1.2) is a probability density function (pdf) in relation to a measure ν . This kernel $K_{\mathbf{x}, \mathbf{H}}(\cdot)$ can be defined as follows.

Definition 2.1. Let $\mathbb{T}_d (\subseteq \mathbb{R}^d)$ be the support of the regressors, $\mathbf{x} \in \mathbb{T}_d$ a target vector and \mathbf{H} a bandwidth matrix. A parametrized pdf $K_{\mathbf{x}, \mathbf{H}}(\cdot)$ of support $\mathbb{S}_{\mathbf{x}, \mathbf{H}} (\subseteq \mathbb{R}^d)$ is called “multivariate (or general) associated kernel” if the following conditions are satisfied:

$$\mathbf{x} \in \mathbb{S}_{\mathbf{x}, \mathbf{H}}, \quad (2.1)$$

$$\mathbb{E}(\mathcal{Z}_{\mathbf{x}, \mathbf{H}}) = \mathbf{x} + \mathbf{a}(\mathbf{x}, \mathbf{H}), \quad (2.2)$$

$$\text{Cov}(\mathcal{Z}_{\mathbf{x}, \mathbf{H}}) = \mathbf{B}(\mathbf{x}, \mathbf{H}), \quad (2.3)$$

where $\mathcal{Z}_{\mathbf{x}, \mathbf{H}}$ denotes the random vector with pdf $K_{\mathbf{x}, \mathbf{H}}$ and both $\mathbf{a}(\mathbf{x}, \mathbf{H}) = (a_1(\mathbf{x}, \mathbf{H}), \dots, a_d(\mathbf{x}, \mathbf{H}))^\top$ and $\mathbf{B}(\mathbf{x}, \mathbf{H}) = (b_{ij}(\mathbf{x}, \mathbf{H}))_{i,j=1,\dots,d}$ tend, respectively, to the null vector $\mathbf{0}$ and the null matrix $\mathbf{0}_d$ as \mathbf{H} goes to $\mathbf{0}_d$.

From this definition and in comparison with (1.2), the Nadaraya-Watson estimator using associated kernels is

$$\widetilde{m}_n(\mathbf{x}; K_{\mathbf{x}, \mathbf{H}}) = \sum_{i=1}^n \frac{Y_i K_{\mathbf{x}, \mathbf{H}}(\mathbf{X}_i)}{\sum_{i=1}^n K_{\mathbf{x}, \mathbf{H}}(\mathbf{X}_i)} = \widetilde{m}_n(\mathbf{x}; \kappa, \mathbf{H}), \quad \forall \mathbf{x} \in \mathbb{T}_d \subseteq \mathbb{R}^d, \quad (2.4)$$

where $\mathbf{H} \equiv \mathbf{H}_n$ is the bandwidth matrix such that $\mathbf{H}_n \rightarrow \mathbf{0}$ as $n \rightarrow \infty$, and κ represents the type of associated kernel $K_{\mathbf{x}, \mathbf{H}}$, parametrized by \mathbf{x} and \mathbf{H} . Without loss of generality and to point out the effect of κ , we will in hereafter use

$\widetilde{m}_n(\mathbf{x}; \boldsymbol{\kappa}, \mathbf{H}) = \widetilde{m}_n(\mathbf{x}; \boldsymbol{\kappa})$ since the bandwidth matrix is here investigated only by cross validation.

The following two examples provide the well-known and also interesting particular cases of multivariate associated kernel estimators. The first can be seen as an interpretation of classical associated kernels through continuous symmetric kernels. The second deals on non-classical associated kernels without correlation structure.

Given a target vector $\mathbf{x} \in \mathbb{R}^d =: \mathbb{T}_d$ and a bandwidth matrix \mathbf{H} , it follows that the classical kernel in (1.2) with null mean vector and covariance matrix $\boldsymbol{\Sigma}$ induces the so-called (multivariate) classical associated kernel:

$$(i) \quad K_{\mathbf{x}, \mathbf{H}}(\cdot) = \frac{1}{\det \mathbf{H}} K\{\mathbf{H}^{-1}(\mathbf{x} - \cdot)\} \quad (2.5)$$

on $\mathbb{S}_{\mathbf{x}, \mathbf{H}} = \mathbf{x} - \mathbf{H}\mathbb{S}_d$ with $\mathbb{E}(\mathcal{Z}_{\mathbf{x}, \mathbf{H}}) = \mathbf{x}$ (i.e. $\mathbf{a}(\mathbf{x}, \mathbf{H}) = \mathbf{0}$) and $\text{Cov}(\mathcal{Z}_{\mathbf{x}, \mathbf{H}}) = \mathbf{H}\boldsymbol{\Sigma}\mathbf{H}$;

$$(ii) \quad K_{\mathbf{x}, \mathbf{H}}(\cdot) = \frac{1}{(\det \mathbf{H})^{1/2}} K\{\mathbf{H}^{-1/2}(\mathbf{x} - \cdot)\}$$

on $\mathbb{S}_{\mathbf{x}, \mathbf{H}} = \mathbf{x} - \mathbf{H}^{1/2}\mathbb{S}_d$ with $\mathbb{E}(\mathcal{Z}_{\mathbf{x}, \mathbf{H}}) = \mathbf{x}$ (i.e. $\mathbf{a}(\mathbf{x}, \mathbf{H}) = \mathbf{0}$) and $\text{Cov}(\mathcal{Z}_{\mathbf{x}, \mathbf{H}}) = \mathbf{H}^{1/2}\boldsymbol{\Sigma}\mathbf{H}^{1/2}$.

A second particular case of Definition 2.1, appropriate for both continuous and count explanatory variables without correlation structure is presented as follows.

Let $\mathbf{x} = (x_1, \dots, x_d)^\top \in \times_{j=1}^d \mathbb{T}_1^{[j]} =: \mathbb{T}_d$ and $\mathbf{H} = \mathbf{Diag}(h_{11}, \dots, h_{dd})$ with $h_{jj} > 0$. Let $K_{x_j, h_{jj}}^{[j]}$ be a (discrete or continuous) univariate associated kernel (see Definition 2.1 for $d = 1$) with its corresponding random variable $\mathcal{Z}_{x_j, h_{jj}}^{[j]}$ on $\mathbb{S}_{x_j, h_{jj}} (\subseteq \mathbb{R})$ for all $j = 1, \dots, d$. Then, the multiple associated kernel is also a multivariate associated kernel:

$$K_{\mathbf{x}, \mathbf{H}}(\cdot) = \prod_{j=1}^d K_{x_j, h_{jj}}^{[j]}(\cdot) \quad (2.6)$$

on $\mathbb{S}_{\mathbf{x}, \mathbf{H}} = \times_{j=1}^d \mathbb{S}_{x_j, h_{jj}}$ with $\mathbb{E}(\mathcal{Z}_{\mathbf{x}, \mathbf{H}}) = (x_1 + a_1(x_1, h_{11}), \dots, x_d + a_d(x_d, h_{dd}))^\top$ and $\text{Cov}(\mathcal{Z}_{\mathbf{x}, \mathbf{H}}) = \mathbf{Diag}(b_{jj}(x_j, h_{jj}))_{j=1, \dots, d}$. In other words, the random variables $\mathcal{Z}_{x_j, h_{jj}}^{[j]}$ are independent components of the random vector $\mathcal{Z}_{\mathbf{x}, \mathbf{H}}$.

Here, in addition to the Nadaraya-Watson estimator using general associated kernels given in (2.4), we proposed a slight one. In fact, for multivariate supports composed of continuous and discrete univariate support, we lack appropriate general associated kernels. Therefore, the estimator (2.4) becomes with multiple associated kernels (2.6):

$$\widetilde{m}_n(\mathbf{x}; \boldsymbol{\kappa}) = \sum_{i=1}^n \frac{Y_i \prod_{j=1}^d K_{x_j, h_{jj}}^{[j]}(X_{ij})}{\sum_{i=1}^n \prod_{j=1}^d K_{x_j, h_{jj}}^{[j]}(X_{ij})}, \quad \forall \mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{T}_d := \times_{j=1}^d \mathbb{T}_1^{[j]}. \quad (2.7)$$

In theory and in practice, one often uses (2.7) from multiple associated kernels (2.6) which are more manageable than (2.4); see, e.g., [Scott \(1992\)](#) and also [Bouerzmarni and Rombouts \(2009\)](#) for density estimation.

2.2. Associated kernels for illustration

In order to point out the importance of the type of kernel κ in a regression study, we motivate below some kernels that will be used in simulations. These concern seven basic associated kernels for which three of them are univariate discrete, three others are univariate continuous and the last one is a bivariate beta with correlation structure.

- The binomial kernel (Bin) is defined on the support $\mathbb{S}_x = \{0, 1, \dots, x+1\}$ with $x \in \mathbb{T}_1 := \mathbb{N} = \{0, 1, \dots\}$ and then $h \in (0, 1]$:

$$B_{x,h}(u) = \frac{(x+1)!}{u!(x+1-u)!} \left(\frac{x+h}{x+1} \right)^u \left(\frac{1-h}{x+1} \right)^{x+1-u} \mathbb{1}_{\mathbb{S}_x}(u),$$

where $\mathbb{1}_A$ denote the indicator function of any given event A . Note that $B_{x,h}$ is the probability mass function (pmf) of the binomial distribution $\mathcal{B}(x+1; (x+h)/(x+1))$ with its number of trials $x+1$ and its success probability in each trial $(x+h)/(x+1)$. It is appropriated for count data with small or moderate sample sizes and, also, it does not satisfy (2.3); see [Kokonendji and Senga Kiéssé \(2011\)](#) and also [Zougab et al. \(2012\)](#) for a bandwidth selection by Bayesian method.

- For fixed arm $a \in \mathbb{N}$, the discrete triangular kernel (DTr a) is defined on $\mathbb{S}_{x,a} = \{x, x \pm 1, \dots, x \pm a\}$ with $x \in \mathbb{T}_1 = \mathbb{N}$:

$$DT_{x,h;a}(u) = \frac{(a+1)^h - |u-a|^h}{P(a,h)} \mathbb{1}_{\mathbb{S}_{x\setminus\{a\}}}(u),$$

where $h > 0$ and $P(a,h) = (2a+1)(a+1) - 2 \sum_{k=0}^a k^h$ is the normalizing constant. It is symmetric around the target x , satisfying Definition 2.1 and suitable for count variables; see [Kokonendji et al. \(2007\)](#) and also [Kokonendji and Zocchi \(2010\)](#) for an asymmetric version.

- From [Aitchison and Aitken \(1976\)](#), [Kokonendji and Senga Kiéssé \(2011\)](#) deduced the following discrete kernel that we here label DiracDU (DirDU) as “Dirac Discrete Uniform”. For fixed $c \in \{2, 3, \dots\}$ the number of categories, we define $\mathbb{S}_{x,c} = \{0, 1, \dots, c-1\}$ and

$$DU_{x,h;c}(u) = (1-h) \mathbb{1}_{\{x\}}(u) + \frac{h}{c-1} \mathbb{1}_{\mathbb{S}_{x,c} \setminus \{x\}}(u),$$

where $h \in (0, 1]$ and $x \in \mathbb{T}_1$. This DiracDU kernel is symmetric around the target, satisfying Definition 2.1 and appropriated for categorical set \mathbb{T}_1 . See, e.g., [Racine and Li \(2004\)](#) for some uses.

- From the well known [Epanechnikov \(1969\)](#) kernel $K^E(u) = \frac{3}{4}(1 - u^2)\mathbb{1}_{[-1,1]}(u)$, we define its associated version (Epan) on $\mathbb{S}_{x,h} = [x - h, x + h]$ with $x \in \mathbb{T}_1 := \mathbb{R}$ and $h > 0$:

$$K_{x,h}^E(u) = \frac{3}{4h} \left\{ 1 - \left(\frac{u - x}{h} \right)^2 \right\} \mathbb{1}_{[x-h, x+h]}(u).$$

It is obtained through (2.5) and is well adapted for continuous variables with unbounded supports.

- The gamma kernel (Gamma) is defined on $\mathbb{S}_{x,h} = [0, \infty) = \mathbb{T}_1$ with $x \in \mathbb{T}_1$ and $h > 0$:

$$GA_{x,h}(u) = \frac{u^{x/h}}{\Gamma(1 + x/h) h^{1+x/h}} \exp\left(-\frac{u}{h}\right) \mathbb{1}_{[0,\infty)}(u),$$

where $\Gamma(\cdot)$ is the classical gamma function. It is the pdf of the gamma distribution $\mathcal{Ga}(1 + x/h, h)$ with scale parameter $1 + x/h$ and shape parameter h . It satisfies Definition 2.1 and suits for non-negative real set \mathbb{T}_1 ; see [Chen \(2000a\)](#).

- The beta kernel (Beta) is however defined on $\mathbb{S}_{x,h} = [0, 1] = \mathbb{T}_1$ with $x \in \mathbb{T}_1$ and $h > 0$:

$$BE_{x,h}(u) = \frac{u^{x/h}(1 - u)^{(1-x)/h}}{\mathcal{B}(1 + x/h, 1 + (1 - x)/h)} \mathbb{1}_{[0,1]}(u),$$

where $\mathcal{B}(r, s) = \int_0^1 t^{r-1}(1 - t)^{s-1} dt$ is the usual beta function with $r > 0$ and $s > 0$. It is the pdf of the beta distribution $\mathcal{Be}(1 + x/h, (1 - x)/h)$ with shape parameters $1 + x/h$ and $(1 - x)/h$. This pdf satisfies Definition 2.1 and is appropriated for rates, proportions and percentages dataset \mathbb{T}_1 ; see [Chen \(1999\)](#).

- We finally consider the bivariate beta kernel (Bivariate beta) defined by

$$\begin{aligned} BS_{\mathbf{x}, \mathbf{H}}(u_1, u_2) &= \left(\frac{u_1^{x_1/h_{11}}(1 - u_1)^{(1-x_1)/h_{11}}}{\mathcal{B}(1 + x_1/h_{11}, 1 + (1 - x_1)/h_{11})} \right) \left(\frac{u_2^{x_2/h_{22}}(1 - u_2)^{(1-x_2)/h_{22}}}{\mathcal{B}(1 + x_2/h_{22}, 1 + (1 - x_2)/h_{22})} \right) \\ &\times \left(1 + h_{12} \times \frac{u_1 - \tilde{\mu}_1(x_1, h_{11})}{h_{11}^{1/2} \tilde{\sigma}_1(x_1, h_{11})} \times \frac{u_2 - \tilde{\mu}_2(x_2, h_{22})}{h_{22}^{1/2} \tilde{\sigma}_2(x_2, h_{22})} \right) \mathbb{1}_{[0,1]^2}(u_1, u_2), \quad (2.8) \end{aligned}$$

with $\mathbb{S}_{\mathbf{x}, \mathbf{H}} = \mathbb{T}_2 = [0, 1]^2$, $\mathbf{x} = (x_1, x_2)^\top \in \mathbb{T}_2$ and $\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{pmatrix}$. For $j = 1, 2$, the characteristics in (2.8) are given by $h_{jj} > 0$, $\tilde{\mu}_j(x_j, h_{jj}) = (x_j + h_{jj})/(1 + 2h_{jj})$, $\tilde{\sigma}_j^2(x_j, h_{jj}) = (x_j + h_{jj})(1 + h_{jj} - x_j)(1 + 2h_{jj})^{-2}(1 + 3h_{jj})^{-1}h_{jj}$, and the constraints

$$h_{12} \in [-\beta, \beta'] \cap \left(-\sqrt{h_{11}h_{22}}, \sqrt{h_{11}h_{22}} \right) \quad (2.9)$$

$$\text{with } \beta = \left(\max_{v_1, v_2} \left\{ \frac{v_1 - \tilde{\mu}_1(x_1, h_{11})}{h_{11}^{1/2} \tilde{\sigma}_1(x_1, h_{11})} \times \frac{v_2 - \tilde{\mu}_2(x_2, h_{22})}{h_{22}^{1/2} \tilde{\sigma}_2(x_2, h_{22})} \right\} \right)^{-1} \text{ and}$$

$$\beta' = \left| \left(\min_{v_1, v_2} \left\{ \frac{v_1 - \tilde{\mu}_1(x_1, h_{11})}{h_{11}^{1/2} \tilde{\sigma}_1(x_1, h_{11})} \times \frac{v_2 - \tilde{\mu}_2(x_2, h_{22})}{h_{22}^{1/2} \tilde{\sigma}_2(x_2, h_{22})} \right\} \right)^{-1} \right|.$$

It satisfies Definition 2.1 and is adapted for bivariate rates. The full bandwidth matrix \mathbf{H} allows any orientation of the kernel. Therefore, it can reach any point of the space which might be inaccessible with diagonal matrix. This type of kernel is called beta-Sarmanov kernel by Kokonendji and Somé (2015); see Sarmanov (1966) and also Lee (1996) for this construction of multivariate densities with correlation structure from independent components. Like Bertin and Klutnitchkoff (2014), the miminax properties of this bivariate beta kernel are also possible and more generally for associated kernels.

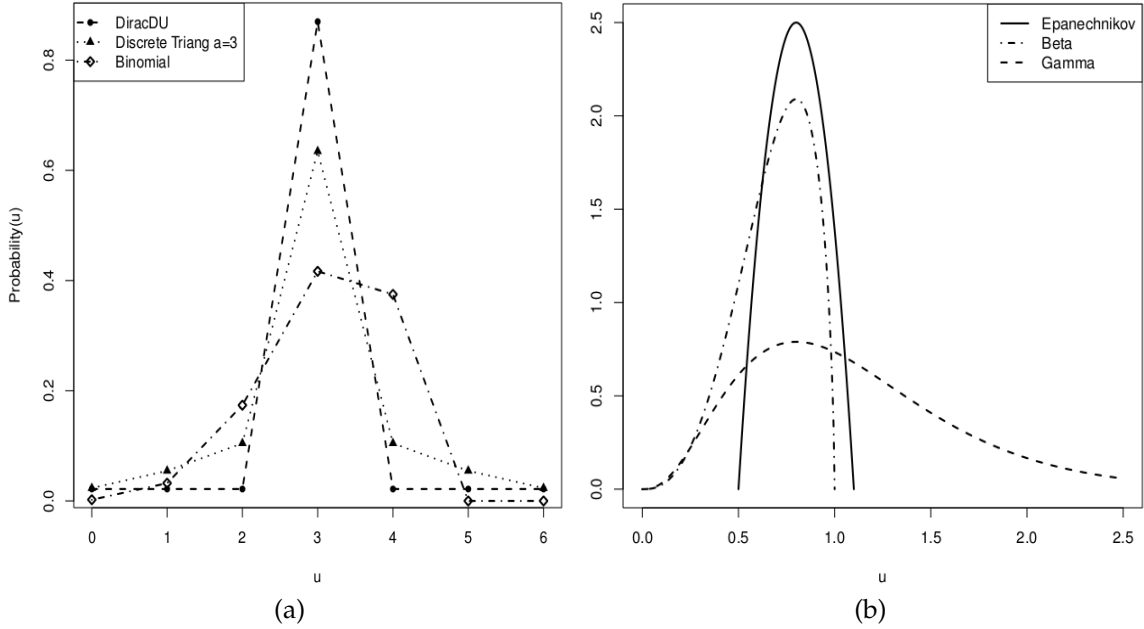


Figure 2.1: Shapes of univariate (discrete and continuous) associated kernels: (a) DiracDU, discrete triangular $a = 3$ and binomial with same target $x = 4$ and bandwidth $h = 0.13$; (b) Epanechnikov, beta and gamma with same $x = 0.8$ and $h = 0.3$.

Figure 2.1 shows some forms of the above-mentioned univariate associated kernels. The plots highlight the importance given to the target point and around it in both discrete and continuous cases. Furthermore, for a fixed bandwidth h , the classical associated kernel of Epanechnikov, and also the categorical DiracDU kernel, keep their respective same shapes along the support; however, they change

according to the target for the others non-classical associated kernels. This explains the inappropriateness of the Epanechnikov kernel for density or regression estimation in any bounded interval (Figure 2.1(a)) and of the DiracDU kernel for count regression estimation (see simulations below).

2.3. Bandwidth matrix selection by cross validation

In the context of multivariate kernel regression, the bandwidth matrix selection is here obtained by the well-known least squares cross-validation. In fact, for a given associated kernel, the optimal bandwidth matrix is $\widehat{\mathbf{H}} = \arg \min_{\mathbf{H} \in \mathcal{H}} \text{LSCV}(\mathbf{H})$ with

$$\text{LSCV}(\mathbf{H}) = \frac{1}{n} \sum_{i=1}^n \{Y_i - \widetilde{m}_{-i}(\mathbf{X}_i; \boldsymbol{\kappa})\}^2, \quad (2.10)$$

where $\widetilde{m}_{-i}(\mathbf{X}_i; \boldsymbol{\kappa})$ is computed as \widetilde{m}_n of (2.4) excluding \mathbf{X}_i and, \mathcal{H} is the set of bandwidth matrices \mathbf{H} ; see, e.g., Kokonendji et al. (2009) in univariate case and also Zhang et al. (2014) and Zougab et al. (2014a) for univariate bandwidth estimation by sampling algorithm methods. For diagonal bandwidth matrices (i.e. multiple associated kernels) the LSCV method use the set of diagonal matrices \mathcal{D} . Concerning the beta-Sarmanov kernel (2.8) with full bandwidth matrix, this LSCV method is used under \mathcal{H}_1 , a subset of \mathcal{H} verifying the constraint (2.9) of the associated kernel. Their algorithms are described below and used for numerical studies in the following section.

Algorithms of LSCV method (2.10) for some type of associated kernels and their corresponding bandwidth matrices

A1. Bivariate beta (2.8) with full bandwidth matrices and dimension $d = 2$.

1. Choose two intervals H_{11} and H_{22} related to h_{11} and h_{22} , respectively.
2. For $\delta = 1, \dots, \ell(H_{11})$ and $\gamma = 1, \dots, \ell(H_{22})$,
 - (a) Compute the interval $H_{12}[\delta, \gamma]$ related to h_{12} from constraints in (2.9);
 - (b) For $\lambda = 1, \dots, \ell(H_{12}[\delta, \gamma])$,
Compose the full bandwidth matrix $\mathbf{H}(\delta, \gamma, \lambda) := (h_{ij}(\delta, \gamma, \lambda))_{i,j=1,2}$ with
 $h_{11}(\delta, \gamma, \lambda) = H_{11}(\delta)$, $h_{22}(\delta, \gamma, \lambda) = H_{22}(\gamma)$ and $h_{12}(\delta, \gamma, \lambda) = H_{12}[\delta, \gamma](\lambda)$.
3. Apply LSCV method on the set \mathcal{H}_1 of all full bandwidth matrices $\mathbf{H}(\delta, \gamma, \lambda)$.

A2. Multiple associated kernels (i.e. diagonal bandwidth matrices) for $d \geq 2$.

1. Choose two intervals H_{11}, \dots, H_{dd} related to h_{11}, \dots, h_{dd} , respectively.
2. For $\delta_1 = 1, \dots, \ell(H_{11}), \dots, \delta_d = 1, \dots, \ell(H_{dd})$,
Compose the diagonal bandwidth matrix $\mathbf{H}(\delta_1, \dots, \delta_d) := \mathbf{Diag}(H_{11}(\delta_1), \dots, H_{dd}(\delta_d))$.

3. Apply LSCV method on the set \mathcal{D} of all diagonal bandwidth matrices $\mathbf{H}(\delta_1, \dots, \delta_d)$.

For a given interval I , the notation $\ell(I)$ is the total number of subdivisions of I and $I(\eta)$ denotes the real value at the subdivision η of I . Also, for practical uses of (A1) and (A2), the intervals H_{11}, \dots, H_{dd} are taken generally according to the chosen associated kernel.

3. Simulation studies and real data analysis

We apply the multivariate associated kernel estimators \tilde{m}_n of (2.4) and (2.7) to some simulated target regressions functions m and then to two real datasets. The multivariate and multiple associated kernels used are built from those of Section 2.2. The optimal bandwidth matrix is here chosen by LSCV method (2.10) using Algorithms A1 and A2 of Section 2.3 and their indications. Besides the criterion of kernel support, we retain three measures to examine the effect of different associated kernels κ on multiple regression. In simulations, it is the average squared errors (ASE) defined as

$$ASE(\kappa) = \frac{1}{n} \sum_{i=1}^n \{m(\mathbf{x}_i) - \tilde{m}_n(\mathbf{x}_i; \kappa)\}^2.$$

For real datasets, we use the root mean squared error (RMSE) which linked to ASE through squared root and by changing the simulated value $m(x_i)$ into the observed value y_i :

$$RMSE(\kappa) = \sqrt{\frac{1}{n} \sum_{i=1}^n \{y_i - \tilde{m}_n(\mathbf{x}_i; \kappa)\}^2}.$$

Also, we consider the practical coefficient of determination R^2 which quantifies the proportion of variation of the response variable Y_i explained by the non-intercept regressor \mathbf{x}_i

$$R^2(\kappa) = \frac{\sum_{i=1}^n \{\tilde{m}_n(\mathbf{x}_i; \kappa) - \bar{y}\}^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

with $\bar{y} = n^{-1}(y_1 + \dots + y_n)$. All these criteria above have their simulated or real data counterparts by replacing y_i with $m(\mathbf{x}_i)$ and vice versa. Computations have been performed on the supercomputer facilities of the Mésocentre de calcul de Franche-Comté using the R software; see [R Development Core Team \(2014\)](#).

3.1. Simulation studies

Expect as otherwise, each result is obtained with the number of replications $N_{sim} = 100$.

3.1.1. Bivariate cases

We consider seven target regression functions labelled A, B, C, D and E with dimension $d = 2$.

- Function A is a bivariate beta without correlation $\rho(x_1, x_2) = 0$:

$$m(x_1, x_2) = \frac{x_1^{p_1-1}(1-x_1)^{q_1-1}x_2^{p_2-1}(1-x_2)^{q_2-1}}{\mathcal{B}(p_1, q_1)\mathcal{B}(p_2, q_2)} \mathbb{1}_{[0,1]}(x_1)\mathbb{1}_{[0,1]}(x_2),$$

with $(p_1, q_1) = (3, 2)$ and $(p_2, q_2) = (5, 2)$ as parameter values in univariate beta density.

- Function B is the bivariate Dirichlet density

$$m(x_1, x_2) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} x_1^{\alpha_1-1} x_2^{\alpha_2-1} (1-x_1-x_2)^{\alpha_3-1} \mathbb{1}_{\{x_1, x_2 \geq 0, x_1+x_2 \leq 1\}}(x_1, x_2),$$

where $\Gamma(\cdot)$ is the classical gamma function, with parameter values $\alpha_1 = \alpha_2 = 5$, $\alpha_3 = 6$ and, therefore, the moderate value of $\rho(x_1, x_2) = -(\alpha_1\alpha_2)^{1/2}(\alpha_1 + \alpha_3)^{-1/2}(\alpha_2 + \alpha_3)^{-1/2} = -0.454$.

- Function C is a bivariate Poisson with null correlation $\rho(x_1, x_2) = 0$:

$$m(x_1, x_2) = \frac{e^{-5} 2^{x_1} 3^{x_2}}{x_1! x_2!} \mathbb{1}_{\mathbb{N}}(x_1) \mathbb{1}_{\mathbb{N}}(x_2).$$

- Function D is a bivariate Poisson with correlation structure

$$m(x_1, x_2) = e^{-(\theta_1 + \theta_2 + \theta_{12})} \sum_{i=0}^{\min(x_1, x_2)} \frac{\theta_1^{x_1+i} \theta_2^{x_2+i} \theta_{12}^i}{(x_1+i)!(x_2+i)!i!} \mathbb{1}_{\mathbb{N} \times \mathbb{N}}(x_1, x_2),$$

with parameter values $\theta_1 = 2$, $\theta_2 = 3$ and $\theta_{12} = 4$ and, therefore, the moderate value of $\rho(x_1, x_2) = \theta_{12}(\theta_1 + \theta_{12})^{-1/2}(\theta_2 + \theta_{12})^{-1/2} = 0.617$; see, e.g., [Yahav and Shmueli \(2012\)](#).

- Function E is a bivariate beta without correlation $\rho(x_1, x_2) = 0$:

$$m(x_1, x_2) = \frac{x_1^{p_1-1}(1-x_1)^{q_1-1} 3^{x_2}}{e^3 \mathcal{B}(p_1, q_1) x_2!} \mathbb{1}_{[0,1]}(x_1) \mathbb{1}_{\mathbb{N}}(x_2),$$

with $(p_1, q_1) = (3, 3)$.

Table 3.1 presents the execution times needed for computing the LSCV method for both bivariate beta kernels with respect to only one replication of sample sizes $n = 50$ and 100 for the target function A. The computational times of the LSCV

n	Bivariate beta	Beta×Beta
50	276.198	7.551
100	647.255	30.081

Table 3.1: Typical Central Processing Unit (CPU) times (in seconds) for one replication of LSCV method (2.10) by using Algorithms A1 and A2 of Section 2.3.

method for the bivariate beta with correlation structure (2.8) are obviously longer than those without correlation structure. Let us note that for full bandwidth matrices, the execution times become very large when the number of observations is large; however, these CPU times can be considerably reduced by parallelism processing, in particular for the bivariate beta kernel with full LSCV method (2.10). These constraints (2.9) reflect the difficulty for finding the appropriate bandwidth matrix with correlation structure by LSCV method.

	n	Bivariate beta	Beta×Beta	$Epan \times Epan$
A	50	0.4368(0.3754)	0.4266(0.3724)	0.7483(0.2342)
	100	0.1727(0.0664)	0.1952(0.0816)	0.6727(0.1413)
B	50	1.2564(0.5875)	1.4267(0.4024)	1.6675(2.0353)
	100	0.3041(0.1151)	0.3362(0.1042)	1.3975(1.5758)

Table 3.2: Some expected values of $\overline{ASE}(\kappa)$ and their standard errors in parentheses with $N_{sim} = 100$ of some multiple associated kernel regressions for simulated continuous data from functions A with $\rho(x_1, x_2) = 0$ and B with $\rho(x_1, x_2) = -0.454$.

Table 3.2 reports the average $ASE(\kappa)$ which we denote $\overline{ASE}(\kappa)$ for three continuous associated kernels κ with respect to functions A and B and according to sample sizes $n \in \{50, 100\}$. We can see that both beta kernels in dimension $d = 2$ work better than the multiple Epanechnikov kernel for all sample sizes and all correlation structure in the regressors. This reflects the appropriateness of the beta kernels which are suitable to the support of rate regressors. Then, the explanatory variables with correlation structure give larger $\overline{ASE}(\kappa)$ than those without correlation structure. Also, both beta kernels give quite similar results. Furthermore, all $\overline{ASE}(\kappa)$ are better when the sample size increases.

Finally, Tables 3.1 and 3.2 highlight that the use of bivariate beta kernels with correlation structure is not recommend in regression with rates explanatory variables. Thus, we focus on multiple associated kernels for the rest of the simulations studies.

Table 3.3 shows the values $\overline{ASE}(\kappa)$ with respect to five associated kernels κ for sample size $n = 20, 50$ and 100 and count datasets generated from C and D.

	n	DTr2×DTr2	DTr3×DTr3	Bin×Bin	Epan×Epan	DirDU×DirDU
C	20	1.5e-6(2.2e-6)	3.3e-6(4.1e-6)	3.6e-5(9.7e-6)	4.0e-5(3.5e-5)	1.6e-8(1.8e-8)
	50	3.1e-7(6.9e-7)	4.7e-7(9.7e-7)	3.6e-5(7.4e-6)	3.8e-5(2.8e-5)	3.7e-9(2.3e-9)
	100	8.6e-8(1.2e-7)	2.9e-7(3.1e-7)	3.7e-5(4.8e-6)	3.6e-5(2.3e-5)	4.1e-10(3.5e-10)
D	20	2.4e-6(2.8e-6)	4.5e-6(4.9e-6)	7.1e-6(2.6e-6)	4.2e-6(2.5e-6)	2.7e-8(2.1e-8)
	50	2.5e-7(3.4e-7)	1.8e-7(2.5e-7)	8.1e-5(4.3e-6)	5.1e-6(1.2e-6)	4.3e-9(3.2e-9)
	100	2.6e-8(6.2e-8)	4.8e-8(9.5e-8)	9.3e-6(8.2e-7)	7.2e-6(7.8e-7)	5.3e-10(4.6e-10)

Table 3.3: Some expected values of $\overline{ASE}(\kappa)$ and their standard errors in parentheses with $N_{sim} = 100$ of some multiple associated kernel regressions for simulated count data from functions C with $\rho(x_1, x_2) = 0$ and D with $\rho(x_1, x_2) = 0.617$.

Globally, the discrete associated kernels in multiple case perform better than the multiple Epanechnikov kernel for all sample sizes and correlation structure in the regressors. The use of categorical DiracDU kernels gives the best result in term of $\overline{ASE}(\kappa)$ but DiracDU does not suit for these count datasets. Also, the discrete triangular kernels gives the most interesting result with an advantage to the discrete triangular with small arm $a = 2$. This discrete triangular is the best since it concentrates always on the target and a few observations around it; see Figure 2.1(a). The results become much better when the sample size increases. The values $\overline{ASE}(\kappa)$ for regressors with or without correlation structure are comparable; and thus, we can focus on target regression functions without correlation structure for the remaining simulations.

	n	Beta×DTr2	Beta×DTr3	Beta×Bin	Beta×Epan	Beta×DirDU
E	30	3.738(1.883)	1.966(1.382)	3.884(1.298)	6.361(2.134)	0.162(0.201)
	50	3.978(1.404)	2.106(1.119)	3.683(0.833)	7.143(1.732)	0.138(0.171)
	100	3.951(1.052)	1.956(0.806)	3.835(0.834)	7.277(1.574)	0.113(0.147)

Table 3.4: Some expected values ($\times 10^3$) of $\overline{ASE}(\kappa)$ and their standard errors in parentheses with $N_{sim} = 100$ of some multiple associated kernel regressions of simulated mixed data from function E with $\rho(x_1, x_2) = 0$.

Table 3.4 presents the values for sample sizes $n \in \{30, 50, 100\}$ and for five associated kernels κ . The datasets are generated from E and the beta kernel is applied on the continuous rate variable of E. We observe the superiority of the multiple associated kernels using discrete kernels over those defined with the Epanechnikov kernel for all sample sizes. Then, the multiple associated kernel with the categorical DiracDU gives the best $\overline{ASE}(\kappa)$ but it is not appropriate for the count variable of E. Also, the values $\overline{ASE}(\kappa)$ are getting better when the sample

size increases.

From Tables 3.2, 3.3 and 3.4, the importance of the type of associated kernel κ which respect the support of the explanatory variables is proven.

3.1.2. Multivariate cases

Since the appropriate associated kernels perform better than the inappropriate ones, we focus in higher dimension $d > 2$ on regression with only suitable associated kernels. Then, we consider two target regression functions labelled F and G for $d = 3$ and 4 respectively. The formulas of the functions are given below.

- Function F is a 3-variate with null correlation:

$$m(x_1, x_2, x_3) = \frac{x_1^{p_1-1}(1-x_1)^{q_1-1}2^{x_2}3^{x_3}}{e^5 \mathcal{B}(p_1, q_1)x_2!x_3!} \mathbb{1}_{[0,1]}(x_1)\mathbb{1}_{\mathbb{N}}(x_2)\mathbb{1}_{\mathbb{N}}(x_3),$$

with $(p_1, q_1) = (3, 2)$.

- Function G is a 4-variate without correlation:

$$m(x_1, x_2, x_3, x_4) = \frac{x_1^{p_1-1}(1-x_1)^{q_1-1}x_2^{p_2-1}(1-x_2)^{q_2-1}2^{x_3}3^{x_4}}{e^5 \mathcal{B}(p_1, q_1)\mathcal{B}(p_2, q_2)x_3!x_4!} \mathbb{1}_{[0,1]}(x_1)\mathbb{1}_{[0,1]}(x_2)\mathbb{1}_{\mathbb{N}}(x_3)\mathbb{1}_{\mathbb{N}}(x_4),$$

with $(p_1, q_1) = (3, 2)$ and $(p_2, q_2) = (5, 2)$.

n	Beta×DTr3×DTr3	Beta×Bin×DTr3	Beta×Beta×DTr3×DTr3
30	0.2501(0.1264)	0.3038(0.1258)	0.7448(0.5481)
50	0.2381(0.0661)	0.2895(0.0162)	0.6055(0.2291)
100	0.2282(0.0649)	0.2822(0.0608)	0.5012(0.2166)

Table 3.5: Some expected values ($\times 10^3$) of $\overline{ASE}(\kappa)$ and their standard errors in parentheses with $N_{sim} = 100$ of some multiple associated kernel regressions of simulated mixed data from 3-variate F and 4-variate G.

Table 3.5 presents the regression study for dimension $d = 3$ and 4 with respect to functions F and G and for sample size $n \in \{30, 50, 100\}$. The values $\overline{ASE}(\kappa)$ show the superiority of the multiple associated kernels using the discrete triangular kernel with $a = 3$ over the one with the binomial kernel. Some results with respect to function G for an associated kernel κ composed by two beta and two discrete triangular kernels with $a = 3$ are also provided. The errors become smaller when the sample size increases.

3.2. Real data analysis

The dataset consists on a sample of 38 family economies from a US large city and is available as the *FoodExpenditure* object in the *betareg* package of [Cribari-Neto and Zeilis \(2010\)](#). The dataset in its current form gives not available (NA) responses for associated kernel regressions especially when we use the discrete triangular or the DiracDU kernel. Then, we extend the original *FoodExpenditure* dataset with its first 20 observations which guarantees some results for the regression, and thus $n = 58$. The dependent variable is *food/income*, the proportion of household *income* spent on *food*. Two explanatory variables are available: the previously mentioned household *income* (x_1) and the *number of residents* (x_2) living in the household with $\widehat{\rho}(x_1, x_2) = 0.028$. We use the Gamma or the Epanechnikov kernel for the continuous variable *income* and the discrete (of Figure 2.1(a)) or the Epanechnikov for the count variable *number of residents*.

The results of the multiple associated kernels for regression are divided in two in Table 3.6. The appropriate associated kernels which strictly follow the support of each variable give comparable results in terms of both $\text{RMSE}(\kappa)$ and $R^2(\kappa)$. In fact, the associated kernels that use the discrete triangular with arm $a = 2$ and 3 give some $R^2(\kappa)$ approximately equal to 64%. The inappropriate kernels give various results. The multiple Epanechnikov kernel and the type of kernel with DiracDU give $R^2(\kappa)$ higher than 80% while the Gamma×Epanechnikov gives $R^2(\kappa)$ less than 50%. Then, a little difference in terms of $\text{RMSE}(\kappa)$ can induce a high incidence on the $R^2(\kappa)$.

	Gamma×DTr2	Gamma×DTr3	Gamma×Bin
Appropriate	0.01409	0.01426	0.01730
	64.2681	64.2708	56.1091
	Epan×Epan	Gamma×Epan	Gamma×DirDU
Inappropriate	0.01451	0.03266	0.01278
	86.0011	47.0181	89.3462

Table 3.6: Some expected values of $\text{RMSE}(\kappa)$ and in percentages $R^2(\kappa)$ of some multiple associated kernel regressions for the *FoodExpenditure* dataset with $n = 58$.

Table 3.7 of the second dataset aims to explain the turnover of a large company by two proportions explanatory variables obtained by survey. The first variable x_1 is the rate of people who like the company and the second one x_2 is the percentage of people who like the strong product of this company. The dataset is obtained in 80 branch of this company. Obviously, there is a significant correlation between these explanatory variables: $\widehat{\rho}(x_1, x_2) = -0.6949$.

Table 3.8 presents the results for the nonparametric regressions with three associated kernels κ . Both beta kernels offer the most interesting results with $R^2(\kappa)$

x_{1i}	x_{2i}	y_i	x_{1i}	x_{2i}	y_i	x_{1i}	x_{2i}	y_i
68.1	54.6	0.8	80.3	9.1	1.5	78.6	31.0	1.6
60.8	4.4	1.3	23.1	83.1	2.3	44.9	3.2	1.0
34.4	36.2	1.2	16.9	90.4	2.0	78.2	13.9	1.8
59.4	27.5	1.3	9.4	79.2	2.8	60.2	35.2	1.1
4.7	81.0	2.9	55.8	21.9	1.3	65.6	26.1	1.6
19.9	97.4	1.2	27.5	75.0	2.2	74.4	12.6	1.6
20.6	73.6	2.4	59.1	12.9	1.4	83.5	13.3	1.8
16.4	42.9	1.1	2.7	93.9	2.4	10.9	83.5	2.6
29.9	74.4	2.0	13.9	56.9	1.4	27.0	77.1	2.2
84.8	26.6	1.6	14.0	92.9	2.1	3.1	67.0	2.2
46.1	66.9	1.2	22.9	43.9	1.1	14.8	72.9	2.5
10.2	86.3	2.5	53.8	56.2	1.0	80.6	16.5	1.6
89.4	32.5	1.6	23.7	61.5	1.5	64.1	28.6	1.5
30.9	46.3	1.1	39.6	67.2	1.4	15.6	90.5	2.0
24.3	37.8	1.2	59.5	45.1	0.9	3.9	68.6	2.5
27.4	74.6	1.9	17.3	81.2	2.6	66.9	43.7	0.9
47.7	61.7	1.1	93.7	28.5	1.5	1.5	65.8	2.3
33.1	83.8	1.5	28.7	82.7	2.0	35.6	43.7	1.0
0.3	83.3	3.0	61.3	70.9	0.6	13.9	25.0	0.8
76.9	35.4	1.2	67.1	24.0	1.7	13.2	70.8	2.2
29.5	44.6	1.3	85.8	36.5	1.2	34.5	73.7	1.8
19.6	67.7	1.9	35.5	76.9	1.8	55.6	6.9	1.3
96.2	26.1	1.7	18.8	55.9	1.3	30.7	9.1	0.9
85.9	28.0	1.5	50.4	17.7	1.4	43.5	15.1	1.0
5.6	39.1	1.1	67.2	8.7	1.5	31.5	36.7	1.2
99.9	7.15	1.3	13.1	59.4	1.7	30.0	21.5	0.8
61.0	31.1	1.4	13.7	75.8	2.5			

Table 3.7: Proportions (in %) of folks who like the company, those who like its strong product and turnover of a company, designed respectively by the variables x_{1i} , x_{2i} and y_i , with $\widehat{\rho}(x_1, x_2) = -0.6949$ and $n = 80$.

approximately equal to 86%. Note that, the multiple Epanechnikov kernel gives lower performance mainly because this continuous unbounded kernel does not suit for these bounded explanatory variables.

4. Summary and final remarks

We have presented associated kernels for nonparametric multiple regression and in presence of a mixture of discrete and continuous explanatory variables; see,

Bivariate beta	Beta×Beta	Epan×Epan
0.10524	0.10523	0.18886
86.6875	86.6874	76.3431

Table 3.8: Some expected values of $\text{RMSE}(\kappa)$ and in percentages $R^2(\kappa)$ of some bivariate associated kernel regressions for turnover dataset in Table 3.7 with $\widehat{\rho}(x_1, x_2) = -0.6949$ and $n = 80$.

e.g., [Zougab et al. \(2014b\)](#) for a choice of the bandwidth matrix by Bayesian methods. Two particular cases including the continuous classical and the multiple (or product of) associated kernels are highlight with the bandwidth matrix selection by cross-validation. Also, six univariate associated kernels and a bivariate beta with correlation structure are presented and used for computational studies.

Simulation experiments and analysis of two real datasets provide insight into the behaviour of the type of associated kernel κ for small and moderate sample sizes. Tables 3.1, 3.2 and 3.8 on bivariate rate regressions can be conceptually summarized as follows. The use of associated kernels with correlation structure is not recommend. In fact, it is time consuming and have the same performance as the multiple beta kernel. Also, these appropriate beta kernels are better than the inappropriate multiple Epanechnikov. For count regressions, the multiple associated kernels built from the binomial and the discrete triangular with small arms are superior to those with the optimal continuous Epanechnikov. Furthermore, the categorical DiracDU kernel gives misleading results since it does not suit for count variables, see Tables 3.3 and 3.4. We advise beta kernels for rates variables and gamma kernels for non-negative dataset for small and moderate sample sizes, and also for all dimension $d \geq 2$; see, e.g., Tables 3.5 and 3.6. Finally, more than the performance of the regression, it is the correct choice of the associated kernel according to the explanatory variables which is the most important. In other words, the criterion for choosing an associated kernel is the support; however, for several kernels matching the support, we use common measures such as the mean integrated squared error. It should be noted that a large coefficient of determination R^2 does not mean good adjustment of the data; see Tables 3.6 and 3.8. Further research on associated kernels for functional regression is conceivable; see, e.g., [Amiri et al. \(2014\)](#) for classical kernels.

References

References

Aitchison, J., Aitken, C.G.G., 1976. Multivariate binary discrimination by the kernel method. *Biometrika* **63**, 413–420.

- Amiri, A., Crambes, C., Thiam, B., 2014. Recursive estimation of nonparametric regression with functional covariate, *Computational Statistics and Data Analysis* **69**, 154–172.
- Bertin, K., Klutchnikoff, N., 2014. Adaptative estimation of a density function using beta kernels, *ESAIM: Probability and Statistics* **18**, 400–417.
- Bouezmarni, T., Rombouts, J.V.K., 2010. Nonparametric density estimation for multivariate bounded data, *Journal of Statistical Planning and Inference* **140**, 139–152.
- Chen, S.X., 1999. A beta kernel estimation for density functions, *Computational Statistics and Data Analysis* **31**, 131–145.
- Chen, S.X., 2000a. Probability density function estimation using gamma kernels, *Annals of the Institute of Statistical Mathematics* **52**, 471–480.
- Chen, S.X., 2000b. Beta kernels smoothers for regression curves, *Statistica Sinica* **52**, 73–91.
- Cribari-Neto, F., Zeilis, A., 2010. Beta regression in R, *Journal of Statistical Software* **34**, 1–24.
- Epanechnikov, V.A., 1969. Nonparametric estimation of a multivariate probability density, *Theory of Probability and Its Applications* **14**, 153–158.
- Hayfield, T., Racine, J.S., 2007. Nonparametric econometrics: the np package, *Journal of Statistical Software* **27**, 1–32.
- Kokonendji, C.C., Senga Kiéssé, T., 2011. Discrete associated kernels method and extensions, *Statistical Methodology* **8**, 497–516.
- Kokonendji, C.C., Senga Kiéssé, T., Demétrio, C.G.B., 2009. Appropriate kernel regression on a count explanatory variable and applications, *Advances and Applications in Statistics* **12**, 99–125.
- Kokonendji, C.C., Senga Kiéssé, T., Zocchi, S.S., 2007. Discrete triangular distributions and nonparametric estimation for probability mass function, *Journal of Nonparametric Statistics* **19**, 241–254.
- Kokonendji, C.C., Somé, S.M., 2015. On multivariate associated kernels for smoothing some density function, [arXiv:1502.01173](https://arxiv.org/abs/1502.01173).
- Kokonendji, C.C., Zocchi, S.S., 2010. Extensions of discrete triangular distribution and boundary bias in kernel estimation for discrete functions, *Statistics and Probability Letters* **80**, 1655–1662.

- Lee, M-L.T., 1996. Properties and applications of the Sarmanov family of bivariate distributions, *Communications in Statistics - Theory and Methods* **25**, 1207–1222.
- Libengué, F.G., 2013. *Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications*. Ph.D. Thesis Manuscript (in French) to Université de Franche-Comté, Besançon, France & Université de Ouagadougou, Burkina Faso, June 2013, **LMB no. 14334**, Besançon.
- Nadaraya, E.A., 1964. On estimating regression, *Theory of Probability and its Applications* **9**, 141–142.
- R Development Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://cran.r-project.org/>.
- Racine, J., Li, Q., 2004. Nonparametric estimation of regression functions with both categorical and continuous data, *Journal of Econometrics* **119**, 99–130.
- Sarmanov, O.V., 1966. Generalized normal correlation and two-dimensionnal Frechet classes, *Doklady (Soviet Mathematics)* **168**, 596–599.
- Scott, W.D., 1992. *Multivariate Density Estimation*, John Wiley and Sons, New York.
- Wansouwé, W.E., Kokonendji, C.C., Kolyang, D.T., 2015. Disake: Discrete associated kernel estimators, URL <http://cran.r-project.org/web/packages/Disake/>.
- Watson, G.S., 1964. Smooth regression analysis, *Sankhya Series A* **26**, 359–372.
- Yahav, I., Shmueli, G., 2012. On generating multivariate Poisson data in management science applications, *Applied Stochastic Models in Business and Industry* **28**, 91–102.
- Zhang, S., 2010. A note on the performance of the gamma kernel estimators at the boundary, *Statistics and Probability Letters* **80**, 548–557.
- Zhang, S., Karunamuni, R.J., 2010. Boundary performance of the beta kernel estimators, *Journal of Nonparametric Statistics* **22**, 81–104.
- Zhang, X., King, M.L., Shang, H.L., 2014. A sampling algorithm for bandwidth estimation in a nonparametric regression model with a flexible error density, *Computational Statistics and Data Analysis* **78**, 218–234.
- Zougab, N., Adjabi, S., Kokonendji, C.C., 2012. Binomial kernel and Bayes local bandwidth in discrete functions estimation, *Journal of Nonparametrics Statistics* **24**, 783–795.

- Zougab, N., Adjabi, S., Kokonendji, C.C., 2014a. Bayesian approach in nonparametric count regression with binomial kernel, *Communications in Statistics - Simulation and Computation* **43**, 1052–1063.
- Zougab, N., Adjabi, S., Kokonendji, C.C., 2014b. Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation, *Computational Statistics and Data Analysis* **75**, 28–38.